sustainability

MDPI

Article

Impact of Big Data on Carbon Emissions: Empirical Evidence from China's National Big Data Comprehensive Pilot Zone

Yali Liu [1], Zhi Li [1], Haonan Chen [1,*] and Xiaoning Cui [2,*]

[1] School of Economics, Capital University of Economics and Business, Beijing 100070, China; 12021030002@cueb.edu.cn (Y.L.); lizhi@cueb.edu.cn (Z.L.)
[2] School of International Business, Beijing Foreign Studies University, Beijing 100086, China
* Correspondence: chn@cueb.edu.cn (H.C.); cuixiaoning@bfsu.edu.cn (X.C.)

Abstract: Big data is a pivotal factor in propelling the digital economy forward and emerges as a novel driver in realizing the goals of carbon peaking and carbon neutrality. This study focuses on a quasi-natural experiment, namely national big data comprehensive pilot zones (NBD-CPZs), and employs a multi-period difference-in-differences (DID) model to identify the influence of big data on carbon emissions. The findings of this study are as follows. Overall, big data significantly reduces carbon emissions within the pilot zones. Mechanism analysis shows that big data reduces urban carbon emissions by promoting green innovation, optimizing energy structure, mitigating capital mismatch and improving public awareness of environmental protection. Heterogeneity analysis shows that the carbon reduction effect of big data are more pronounced in cities with high levels of digital economy, non-resource-based cities, cities with strong intellectual property rights protection and the Guizhou Province. Spatial effect analysis indicates that within a radius of 400–500 km, the NBD-CPZ increases urban carbon emissions, signifying a significant siphoning effect; within a radius of 500–900 km, the NBD-CPZ reduces urban carbon emissions, signifying a significant spillover effect, and beyond a distance of 900 km, the spatial effect of the NBD-CPZ is not significant. Based on the above conclusions, this study puts forward several policy recommendations to effectively exert the carbon emission reduction effect of big data.

Keywords: big data; national big data comprehensive pilot zone; carbon emission; spatial effect

check for updates

Citation: Liu, Y.; Li, Z.; Chen, H.; Cui, X. Impact of Big Data on Carbon Emissions: Empirical Evidence from China's National Big Data Comprehensive Pilot Zone. *Sustainability* **2024**, *16*, 8313. https://doi.org/10.3390/su16198313

Received: 8 August 2024
Revised: 22 September 2024
Accepted: 23 September 2024
Published: 24 September 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In pursuit of carbon emission reduction and climate change mitigation, the Chinese government has made a solemn commitment toward the goal of carbon peaking and carbon neutrality. As a national imperative, this goal embodies distinctive national attributes and contemporaneous relevance. In the era of digital economy, the scale of China's digital economy is expanding rapidly, ranking second in the world. On the one hand, big data serves not only as a catalyst for industrial transformation and economic growth, but also plays a pivotal role in China's low-carbon transition process. The widespread adoption of big data acts as a powerful enabler, facilitating the dissolution of enterprise barriers, fostering the exchange of eco-friendly knowledge and technology, nurturing inter-firm collaborations and refining carbon mitigation strategies, thereby propelling the green transformation of energy-intensive enterprises. Furthermore, big data enables centralized control management and the intelligent analysis of vast datasets, particularly in domains such as carbon footprints and sink. The application of big data in these domains empower enterprises to meticulously monitor and strategize the mitigation of carbon emissions at every stage of production and operation, thereby facilitating tangible reductions in carbon emissions. However, on the other hand, while big data shows the potential for energy saving and emission reduction, its own storage, processing and transmission processes also result in certain carbon emissions. The storage, processing and transmission of big data rely on computer

hardware, computer software and significant electricity consumption, which consume energy and generate carbon emissions. With the increasing popularity of big data applications, the surge in activities associated with data storage, processing and transmission will inevitably exacerbate carbon emissions. According to the estimations by experts at the Environmental Planning Institute of the Ministry of Ecology, carbon emissions stemming from the nation's data centers totaled 135 million tons in 2021, representing approximately 1.14% of China's overall carbon emissions, with projections indicating an increase to 2% by 2035. This illustrates the "carbon effect paradox" of big data, whereby its utilization results in significant carbon emissions even as it facilitates emissions reduction within traditional industries. Against this backdrop, we understand the direction of big data's impact on carbon emissions and elucidating the underlying transmission mechanisms assumes paramount importance and great guiding and theoretical significance for realizing the low-carbon transition and national high-quality developments driven by big data.

In September 2015, in an effort to promote the advancement of big data, the State Council issued the "Outline for the Promotion of Big Data Development" (hereinafter referred to as the Outline), thereby enshrining the elevation of big data development as a national strategic imperative. The Outline highlighted the developmental advantages and potential inherent in China's big data landscape, while also acknowledging various shortcomings, such as inadequate data accessibility, sharing mechanisms and a nascent industrial foundation. Consequently, the Outline advocated for the enhancement of big data development through the implementation of regional pilot initiatives. Subsequently, in February 2016, Guizhou Province became the first recipient of approval for the establishment of a national big data comprehensive pilot zone (NBD-CPZ). Following this, in October of the same year, the regions of Shanghai, Chongqing, Henan, Shenyang, Inner Mongolia, Beijing-Tianjin-Hebei and the Pearl River Delta were also approved to initiate the construction of NBD-CPZs. The construction of NBD-CPZs can promote the overall development of the data factor market, which in turn, through their radiation effect and demonstration effects, will promote quality change, dynamic change and efficiency change across economic and social spheres. Given the current absence of standardized metrics for assessing big data comprehensively, the establishment of NBD-CPZs, as a pivotal policy endeavor aimed at propelling big data development, furnishes a conducive research conditions for investigating the relationship between big data and carbon emissions.

Since cities are the primary battleground for China's low-carbon economic transformation and high-quality development, this study takes cities as its research object. It empirically explores whether big data has a carbon emission reduction effect, and if so, through what mechanism is it transmitted? Is there any heterogeneity? Is there a spatial effect? To accomplish this, the study empirically examines the impact and spatial effect of big data on carbon emissions using city panel data from 2012 to 2019. It leverages the establishment of NBD-CPZs as a quasi-natural experiment and employs various analytical methods including multi-period difference-in-differences (DID), difference-in-difference propensity score matching (PSM-DID) and spatial difference-in-difference (SDID) models. This study makes three significant contributions to the field. First, this study enriches the related research on the impact of digital technology on carbon emissions. Through a systematic analysis of the "carbon effect paradox" associated with big data, this study provides empirical evidence demonstrating that big data promotes carbon emission reductions, clarifying the paradox and elucidating the environmental implications of big data. Second, this study analyzes the mechanism of big data to improve the environment. By leveraging the exogenous policy of the NBD-CPZs, this study empirically evaluates the environmental impact of big data from a carbon emissions perspective, thereby elucidating the mechanisms through which big data influences carbon emissions and aiding in identifying the focal points for environmental improvement efforts. Third, this study delves deeper into the spatial effect of the impact of big data on carbon emissions. This study uses the SDID model to empirically analyze the spatial effect of NBD-CPZs on carbon emissions at different geographical distances. Our findings provide pragmatic evidence for optimizing

the arrangement of data center nodes and fostering regional collaboration to effectively reduce carbon emissions.

The remainder of this study is structured as follows. Section 2 summarizes the relevant studies at home and abroad. Section 3 puts forward a research hypotheses based on theoretical analysis. Section 4 provides a detailed description of the model's construction, variable description and data source. In Section 5, we analyze the empirical results. Section 6 further examines the spatial effect of NBD-CPZs. Finally, Section 7 lays out the main conclusions of this study and makes relevant policy recommendations.

2. Literature Review

In the digital economy era, big data has emerged as a novel production factor permeating all facets of societal production and life, serving as a pivotal catalyst for China's economic metamorphosis. Scholars at home and abroad have extensively investigated the multifaceted impacts of big data, predominantly focusing on its economic implications. By releasing technological, institutional and innovative dividends, big data has been instrumental in expediting the digital transformation of enterprises [1,2], fostering green innovation [3,4], facilitating the upgrading and transformation of traditional industries [5,6], and augmenting regional total factor productivity [7,8], thereby fostering high-caliber economic development [9–11]. Concurrently, big data has profoundly influenced employment [12], labor income distribution [13,14] and common prosperity [15]. Furthermore, some scholars have focused on the environmental effects of big data. The rapid development of big data offers a pivotal avenue for promoting green development in China [16]. Cheng and Duan (2023) [17] conducted a rigorous analysis to ascertain the environmental effect of big data from the dual perspective of haze pollution and carbon emissions. Chang et al. (2023) [18] approached the subject from the angle of carbon emissions stemming from electricity consumption, revealing a substantial reduction facilitated by big data. This reduction predominantly stems from the promotion of technological innovation and the advancement of inclusive financial practices. Wei et al. (2024) [19] took 277 cities in China as research samples and employed a dual machine learning model to investigate the impact of big data on urban PM2.5 concentration. Their findings underscored the significant reduction in urban PM2.5 concentration attributable to big data development, with urban development and land use planning playing an important role in this process.

In the context of the digital economy, in order to deal with the challenge of global warming, scholars have extensively researched the relationship between the digital economy and carbon emissions [20,21]. Among them, there are fewer studies focusing on the impact of digital technology on carbon emissions and no unanimous conclusions have been reached, delineating two predominant perspectives: the "promotion theory" and the "inhibition theory". The "promotion theory" posits that the rapid development of digital technology and its associated industries engenders a surge in carbon emissions. For instance, the deployment of Information and Communication Technology (ICT) is often linked to heightened electricity consumption [22]. As one of the main sources of global carbon dioxide, the increase in electricity consumption will inevitably promote carbon dioxide emissions [23]. Conversely, the "inhibition theory" contends that digital technologies can foster reductions in carbon emissions, consequently improving environmental quality [24]. For instance, Lu (2018) [25] scrutinized 12 Asian countries and delineated that ICT can effectively reduce carbon emissions. Shen et al. (2023) [26] utilized panel data of Chinese cities from 2006–2020 to confirm that digital technology plays a pivotal role in diminishing carbon emissions through the facilitation of green innovation and the mitigation of energy intensity.

A thorough examination of existing literature reveals a predominant focus on the economic effects of big data, with relatively scant attention directed towards its environmental effects, and there is an extreme lack of research on the environmental effect of big data from the perspective of carbon emissions. Furthermore, scholarly discourse on the relationship between digital technology and carbon emissions remains inconclusive. Conse-

quently, this study firstly integrates both the direct and indirect effects of big data on carbon emissions into an analytical framework, and theoretically explores the impact of big data on carbon emissions. Subsequently, we empirically investigate the impact of big data on carbon emissions, based on the exogenous policy shock of the construction of NBD-CPZs, employing the multi-period DID model and the PSM-DID model. Finally, the spatial effects of NBD-CPZs are further analyzed through the employment of the SDID model.

3. Theoretical Analysis and Research Hypotheses

3.1. Impact of Big Data on Carbon Emissions

3.1.1. Direct Impact of Big Data on Carbon Emissions

The direct impact of big data on carbon emissions manifests as a carbon-increasing effect. Big data applications entail intricate processes such as data storage, processing and transmission, which are inseparable from the operation of data centers and the utilization of computing resources, necessitating substantial energy input. In addition, as the scope of big data applications expands and the volume of data increases, corresponding expansions and upgrades are imperative for big data systems. The increase in data volume is directly related to heightened energy consumption throughout the processes of data storage, processing and transmission. The majority of this energy is sourced from fossil fuels such as coal and oil, resulting in significant carbon emissions. Thus, the adoption of big data directly contributes to carbon emissions.

3.1.2. Indirect Impact of Big Data on Carbon Emissions

Today, big data is empowering various sectors, and under the pressure of the dual carbon target, big data provides a new impetus for carbon emission reduction. First, big data can reduce urban carbon emissions by promoting green innovation. Big data technology empowering the real economy can reduce information asymmetry [27], enhance the investment confidence of external investors, increase investment in enterprise green innovation projects, mitigate enterprise financing constraints and promote green innovation [28]. Moreover, the utilization of big data technologies enables the government to redress their informational asymmetry, effectuating a shift in environmental governance from reactive to proactive modalities through digital platforms like urban environmental monitoring systems. This facilitates the enhanced oversight of enterprises' pollution activities, thereby compelling them towards green innovation initiatives. Green innovation has a significant inhibiting effect on carbon emissions [29,30]. On the one hand, the wide application of green innovation in enterprise operations and residential life fosters the adoption of clean energy and encourages eco-friendly energy consumption, which simultaneously suppresses carbon emissions at both the production and consumption ends. On the other hand, green technology offers crucial technical support for the treatment of carbon dioxide, thereby realizing the end management of carbon emissions. Thus, it can be seen that big data is conducive to promoting green innovation and, thus, achieving carbon emission reduction.

Second, big data can reduce urban carbon emissions by optimizing energy structures. On the one hand, the empowerment of energy enterprises by big data fosters their digital transformation [2]. This facilitates the establishment of digital energy production platforms and promotes information sharing, thereby accelerating the speed of popularization of new production technologies and realizing the intelligent transformation and upgrading of the production process of energy enterprises [31]. This can effectively alleviate energy overcapacity; furthermore, it promotes the structural refinement of the entire energy sector, thereby contributing to the overarching objective of carbon emission reduction. On the other hand, enterprises can harness big data analytics to conduct comprehensive and methodical evaluations of consumer demands. This can enable them to accurately grasp consumer demand, rationally allocate resources, avoid overproduction by enterprises and thus reduce total energy consumption [32]. Furthermore, big data facilitates inter-regional deployment strategies between conventional and renewable energy sources. This not only furnishes technical backing for refining the cleanliness of energy consumption, but also enables the

optimization of energy consumption patterns. Consequently, this strategic utilization of big data serves as a potent tool in mitigating carbon emissions at their source, thereby optimizing the energy structure and achieving significant reductions in carbon emissions.

Again, big data can reduce carbon emissions by mitigating resource mismatch. Big data can transcend temporal and spatial constraints, thereby facilitating efficacious resource allocation across diverse regions and improving the efficiency of resource allocation. Moreover, the application of big data across diverse industries ensures the acquisition of more timely and accurate information, which enhances market efficacy and reduces search and transaction costs. This can help to dismantle transactional barriers [33] and facilitate the flow of high-quality resources towards high-efficiency sectors, while exerting pressure on inefficient industries to improve their resource allocation efficiency. Big data eases resource mismatch and ultimately empowers industrial structure upgrading [34]. Such upgrades entail a shift from low-level to high-level industrial structures, leading to the gradual aggregation of low-energy-consuming sectors, reduced reliance on fossil fuels, the promotion adoption of clean energy, the facilitation of industrial low-carbon transformations and the eventual reduction of carbon emissions. Thus, it becomes evident that big data contributes to mitigating resource mismatches, promoting industrial advancement and ultimately facilitating carbon emission reduction.

Finally, big data can reduce urban carbon emissions by improving public awareness of environmental protection. On the one hand, big data technology facilitates the public to obtain more real-time and accurate environmental information, which helps the public to understand more intuitively their own environmental conditions, and helps to enhance the public's attention to the environment and then improve public awareness of environmental protection. On the other hand, the government can use big data technology to disseminate knowledge about environmental laws and regulations through social media and other channels, which can help accelerate the popularization and dissemination of environmental knowledge and thus enhance public awareness of environmental protection. Public participation is an important means of environmental governance [35]. Enhanced environmental awareness can enhance the enthusiasm of the public in participating in energy-saving and low-carbon activities, increase the consumption of green products and help promote the formation of green and low-carbon behavior. In addition, with the enhancement of environmental awareness, the public demand for green products increases, which in turn forces enterprises to increase investment in green and low-carbon technologies and promotes the innovation and application of green and low-carbon technologies, which ultimately helps to reduce carbon emissions.

The total impact of big data on carbon emissions hinges on the relative magnitude of its direct and indirect effects. In order to understand the impact of big data on carbon emissions in a more visual and intuitive way, this study examines the direct, indirect and total effects of big data on carbon emissions on the same coordinate axis, as illustrated in Figure 1. Among these, curve h1 depicts the direct effect of big data on carbon emissions, representing its carbon-increasing impact. Given that the direct increase of big data on carbon emissions correlates positively with the energy consumption stemming from data scale expansion, the direct effect of big data follows a linear function. Curve h2 illustrates the indirect carbon reduction effect of big data. As the amount of data continues to increase, the value of data will also increase. When the amount of data reaches a certain scale, quantitative changes will eventually incur qualitative changes, and, thus, the value of data will be characterized by exponential growth. As evidenced in prior analyses, big data can mitigate carbon emissions by promoting green innovation, optimizing energy structures, mitigating resource mismatch and improving public awareness of environmental protection. This means that when the amount of data reaches a certain scale, the impact of big data on green innovation, energy structure, resource mismatch and public awareness of environmental protection will be greatly increased, which will lead to a greater carbon emission reduction effect. Consequently, the indirect effect of big data follows a non-linear trajectory. Curve

h3 portrays the overall effect of big data on carbon emissions, representing the combined effect of the above two effects. In summary, this study posits the following hypothesis:

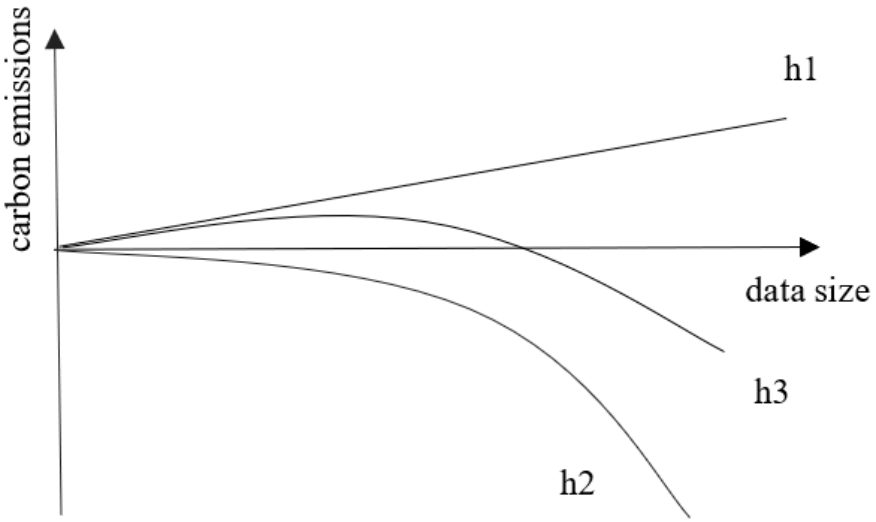


Figure 1. Direct, indirect and total impacts of big data on carbon emissions.

Hypothesis 1a. *When the direct effect outweighs the indirect effect, big data will increase urban carbon emissions.*

Hypothesis 1b. *When the indirect effect outweighs the direct effect, big data will reduce urban carbon emissions.*

Hypothesis 2. *Big data indirectly reduces urban carbon emissions by promoting green innovation, optimizing the energy structure, mitigating resource mismatch and improving public awareness of environmental protection.*

3.2. Space Effects

Considering the existence of an obvious spatial correlation with big data, NBD-CPZs engender spatial effects, principally evidenced through siphon and spillover effects.

NBD-CPZs can generate a siphon effect. The construction of NBD-CPZs prioritizes internal development, aiming to dismantle barriers to data accessibility, fortify infrastructure coordination, foster pivotal big data enterprises and nurture expertise within the big data industry. These initiatives enhance the attractiveness of the pilot zones, leading to the concentration of capital and talent within them. Consequently, neighboring regions may experience a dearth of incentives for green innovation, owing to the outflow of essential resources, thus impeding progress towards carbon emission reduction. Moreover, the establishment of NBD-CPZs instigates data agglomeration, fostering spillover effects predominantly attributed to the positive externalities inherent in data. As data are accumulated and utilized, the scale of the data continues to increase, forming a positive feedback effect. Concurrently, the data factor itself elicits spillover effects. In contrast to traditional factors of production, data is a novel resource characterized by its low cost, widespread accessibility and expansive scale. It can transcend temporal and spatial constraints, alleviate information asymmetry, facilitate unrestricted resource mobility across regions and mitigate resource mismatches, thereby reducing carbon emissions. Furthermore, big data expedites the interregional dissemination, exchange and diffusion of knowledge and technology, optimizes the allocation of spatial resources and thus reduces carbon emissions. Consequently, it becomes evident that NBD-CPZs simultaneously exhibit both siphon and spillover effects.

Furthermore, both the siphon effect and the spillover effect of NBD-CPZs have a functioning boundary. The changing trend of their spatial effect is elucidated in Figure 2,

where curve l1 denotes the siphon effect and curve l2 denotes the spillover effect. Drawing from spatial economics theory, it becomes apparent that the central region acts as a magnet for resources from its periphery, inducing a flow of resources from the surrounding areas towards the core, thereby engendering an agglomeration shadow in the vicinity. Beyond a certain threshold distance, as the agglomeration shadow dissipates, the central region manifests a pronounced spillover effect anew. It is noteworthy that the magnitude of the siphoning effect exerted by big data on neighboring regions diminishes notably with increasing distance, owing to heightened impediments to factor mobility. Upon surpassing the agglomeration shadow, the NBD-CPZ exhibits significant spillover effects. Primarily derived from data, the spillover effect of the NBD-CPZ transcends spatial and temporal confines. While, theoretically, the inter-provincial spillover effect remains homogeneous, practical disparities in data infrastructure, systems and other facets among cities cause spillovers to weaken with geographic distance, resulting in boundaries. As the distance expands beyond the functioning boundary of the spillover effect, the spillover effect of the NBD-CPZ is no longer significant. Curve l3 denotes the total spatial effect, reflecting the relative magnitude of both the siphon effect and the spillover effect. To summarize, this study posits the following hypothesis:

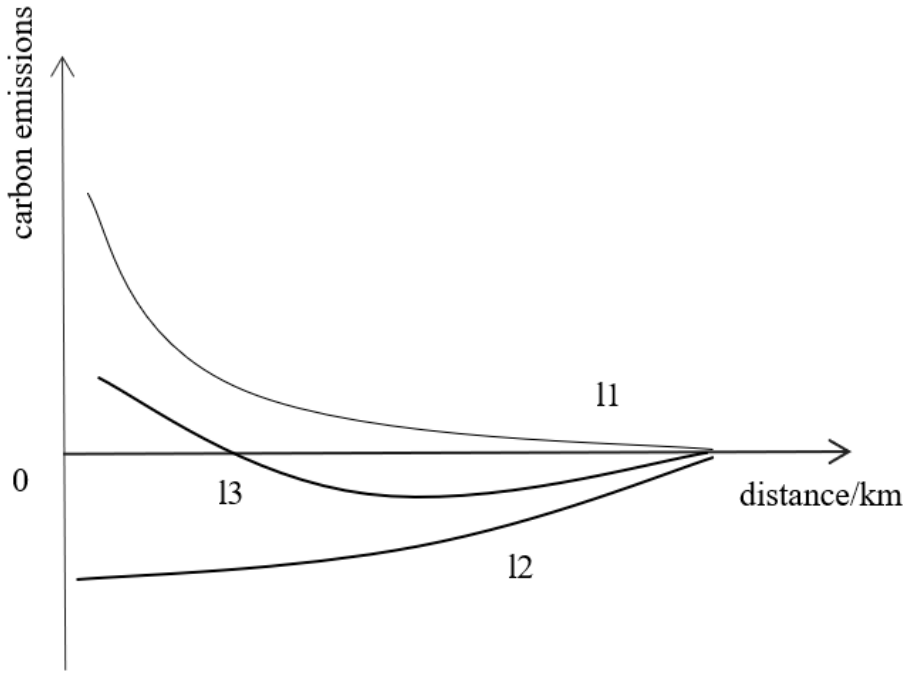


Figure 2. Spatial effects.

Hypothesis 3. *NBD-CPZs have both siphon effects and spillover effects, and there are certain spatial structures.*

4. Research Design

4.1. Methodology

Based on the above theoretical analysis, this study empirically explores the impact of big data on carbon emissions by using the multi-period DID model based on the policy shock of NBD-CPZs, and further adopts the SDID model to examine the spatial effect of the impact of NBD-CPZs on carbon emissions. The methodology framework is visually represented in Figure 3.

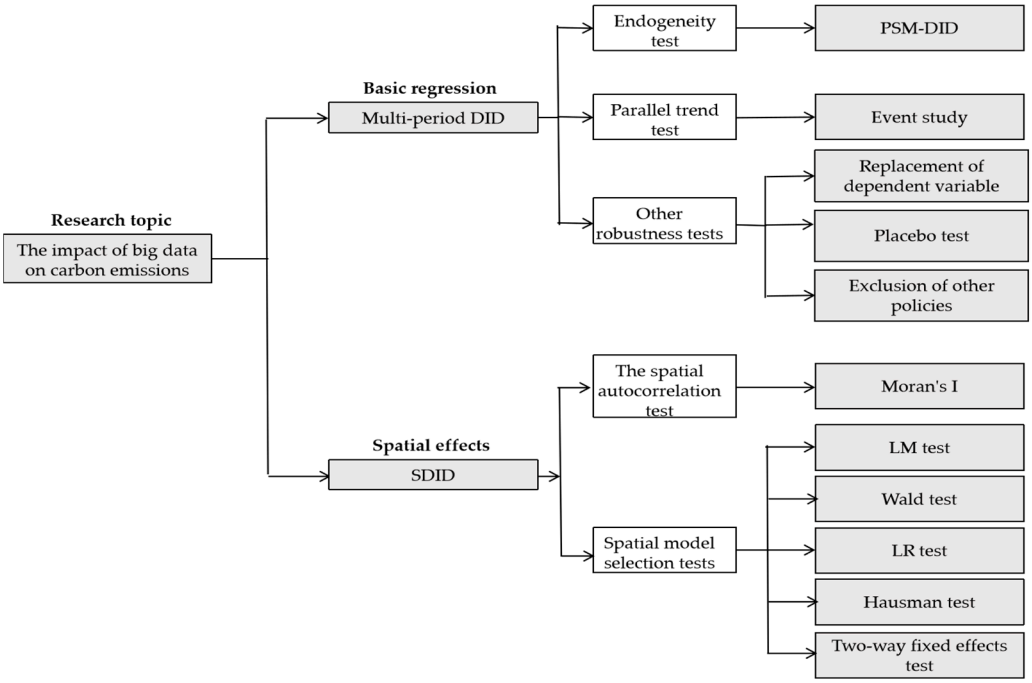


Figure 3. Methodology framework.

Drawing on the research of Beck et al. (2010) [36], this study sets the multi-period DID model as follows:

$$Carbon_{it} = \alpha_0 + \alpha_1 Did_{it} + \gamma X_{it} + \mu_i + \nu_t + \varepsilon_{it} \quad (1)$$

where $Carbon_{it}$ represents the carbon emission level of the city i in the year t; Did_{it} denotes the dummy variable of the NBD-CPZ, with its coefficient α_1 under scrutiny in this study. A significantly negative α_1 would suggest that big data exerts an inhibiting effect on carbon emissions. Additionally, X_{it} signifies the control variable, while μ_i and ν_t denote the fixed effects for city and year, respectively and ε_{it} represents the random perturbation term. Ultimately, the model is a multi-period DID model featuring two-way fixed effects.

4.2. Selection of Variables and Description of Data

4.2.1. Dependent Variable (*Carbon*)

The dependent variable in this study is the level of urban carbon emissions. Currently, there is no official organization to centralize the measurement of carbon emission data at the city level. In light of concerns regarding data comprehensiveness, consistency and comparability, this research adopts the methodology outlined by Zhang and Feng (2020) [37]. This approach involves the conversion of various energy sources such as natural gas, oil, gas, electricity and heat into their corresponding carbon emissions, based on respective emission coefficients. Subsequently, these emissions are aggregated to derive the total carbon emissions for each city, with the logarithm of the total carbon emission level being taken for analysis.

4.2.2. Independent Variable (*Did*)

The independent variable of this study is the policy dummy variable of the NBD-CPZ. Presently, ten provinces and cities across China have received approval for establishing an NBD-CPZ. The policy dummy variable is set according to the approved region and time of the pilot zones. If the city belongs to the approved region and is in its year of approval or later, the dummy variable (*Did*) is set to 1; otherwise, it is set to 0. It is pertinent to highlight that, for Guizhou Province, the policy impact time is 2015, whereas for other pilot cities, the policy impact time is set as 2016.

4.2.3. Control Variables

Based on existing research, this study controls for the following five variables. (1) Urban greening level (*Green*): urban greening is of great significance in urban ecosystems [38,39]. Li and Wang (2023) [40] pointed out that urban greening can absorb a large amount of carbon emissions, which in turn is conducive to reducing urban carbon emissions. In this study, the green coverage of urban built-up areas is used to measure the level of urban greening. (2) The economic development level (*Pgdp*): the rapid development of a city's economy is not without a large amount of energy consumption, which, in turn, will directly lead to an increase in urban carbon emissions. In addition, the implementation of urban carbon emission reduction measures depends on a large amount of financial support; the higher the level of the city's economy, the more it can provide financial support for these measures and thus promote urban carbon emission reduction. In this study, the logarithm of per capita GDP is used to measure the level of urban economic development. (3) Population size (*Popula*): Liu et al. (2021) [41] pointed out that the population size of cities is positively correlated with the total amount of carbon dioxide emissions in China. In addition, other scholars believe that the population size suitable for urban development is conducive to reducing carbon emissions [42]. In this study, we use the logarithm of a city's year-end population to measure the city's population size. (4) The level of informatization (*Inform*): the construction of information and communication infrastructure can promote technological progress [24], promote industrial upgrading [43] and thus reduce carbon emissions. However, the large-scale construction of ICT leads to increased energy consumption, which is not conducive to urban carbon emission reduction [44]. This study uses the proportion of the number of Internet users to the total population of the region to measure the level of informatization. (5) Government intervention (*Gov*): reasonable government intervention can promote the realization of dual-carbon goals [45], but once the government over-intervenes, it is detrimental to carbon emission reduction. This study adopts the proportion of fiscal expenditure to regional GDP to measure government intervention.

4.3. Data Sources and Descriptive Statistics

The data utilized in this study are sourced from the China Urban Statistical Yearbook and the EPS database. To ensure the integrity and consistency of the sample data, exclusions were made for regions such as Hong Kong, Macao, Taiwan and Tibet and cities with significant data gaps. Consequently, a final panel dataset comprising 283 prefecture-level cities was curated, of which 80 cities had established an NBD-CPZ. Given the limited availability of data about natural gas, oil, gas, electricity and heat energy crucial for city-level carbon emission measurement, the study period spans from 2012 to 2019. The descriptive statistics for the aforementioned variables are presented in Table 1.

Table 1. Descriptive statistics of variables.

Variables	N	Mean	SD	Min	Max	Correlation Coefficient
Carbon	2262	6.38	1.12	3.45	9.03	1
Did	2264	0.14	0.35	0	1	0.256 ***
Green	2199	39.88	5.80	19.03	57.02	0.307 ***
Pgdp	2251	10.48	0.66	9.18	12.25	0.641 ***
Popula	2262	5.89	0.68	3.81	7.22	0.303 ***
Inform	2235	0.23	0.17	0.04	0.98	0.553 ***
Gov	2262	0.26	0.14	0.09	0.81	−0.335 ***

Note: *** $p < 0.01$.

5. Empirical Results and Analysis

5.1. Basic Regression

This study employs a stepwise regression approach to investigate the influence of big data on carbon emissions' estimating model (1), with the results delineated in Table 2. Regression (1) is conducted without the inclusion of any control variables, revealing a signif-

icantly negative coefficient for the policy dummy variable at the 1% level. This suggests that the establishment of an NBD-CPZ contributes to carbon emission reduction. Subsequent regressions (2)–(6) involve a gradual addition of control variables. Regardless of the number of variables introduced, the coefficient for the policy dummy variable remains consistently negative. Upon integrating all selected control variables in regression (6), the coefficient for the policy dummy variable stands at −0.077 and is significant at the 5% level. This means that out of one unit of carbon emissions, an NBD-CPZ can reduce carbon emissions by 0.077 units. In conclusion, big data demonstrates a significant capacity for mitigating carbon emissions, with the indirect effect of big data on carbon emissions surpassing the direct effect, thereby substantiating Hypothesis 1b. In addition, although NBD-CPZs have low inhibitory effects on carbon emissions, big data unleashes great potential in promoting carbon emission reduction in the context of long-term policy implementation. Therefore, China should actively promote the construction of an NBD-CPZ, which can not only promote the development of the big data industry, but also help realize its dual carbon goal from the perspective of carbon emission reduction.

Table 2. Impact of big data on carbon emissions.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Did	−0.131 ***	−0.112 **	−0.067 *	−0.066 *	−0.087 **	−0.077 **
	(0.035)	(0.036)	(0.037)	(0.037)	(0.037)	(0.037)
Green		0.007 **	0.006 *	0.006 *	0.006 *	0.005 *
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Pgdp			0.360 ***	0.354 ***	0.323 ***	0.525 ***
			(0.073)	(0.074)	(0.077)	(0.085)
Popula				−0.207	−0.251	−0.021
				(0.158)	(0.159)	(0.163)
Inform					−0.202 *	−0.162
					(0.109)	(0.110)
Gov						1.055 ***
						(0.261)
Constant	5.162 ***	4.888 ***	1.325 *	2.301 **	2.808 **	−0.608
	(0.187)	(0.229)	(0.760)	(1.155)	(1.195)	(1.314)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2262	2199	2189	2189	2162	2162
R-square	0.907	0.910	0.910	0.910	0.910	0.912

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; robust standard errors in parentheses.

5.2. PSM-DID Retest

Given that the selection process for NBD-CPZs is not entirely random, leading to potential bias in the estimation of multi-period DID, this study employs the PSM-DID model to reassess the carbon impact of big data. Among them, the urban carbon emission level serves as the outcome variable, while the urban greening level, economic development level, population size, informatization level and government intervention are designated as covariates. A one-to-one nearest neighbor matching method is utilized for matching. It is imperative to note that the model must undergo both the common value test and the balance test before the PSM-DID reassessment. Figure 4 illustrates the kernel density plot of the propensity scores of the experimental and control groups pre- and post-matching. It is observed that there exists an overlapping segment of propensity scores between the experimental and control groups pre-matching, satisfying the common value assumption. Furthermore, the distribution of propensity scores between the two groups becomes more consistent post-matching, indicating the efficacy of the matching process. Figure 5 presents the standard deviation of both the experimental and control groups before and after matching. It is evident that post-matching, the absolute value of the standard deviation for each covariate between the two groups is less than 10%. This indicates that matching

effectively addresses the characteristic differences between the experimental and control groups, passing the balance test.

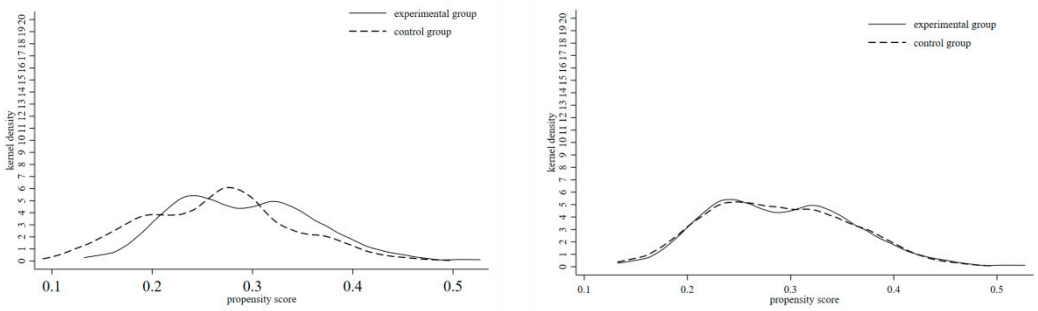


Figure 4. Kernel density plots of propensity scores for the experimental and control groups before (**left** panel) and after (**right** panel) experimental matching.

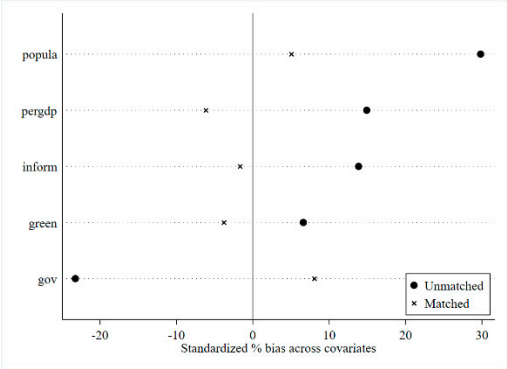


Figure 5. Standardized bias (%) before and after matching in experimental and control groups.

For the matched samples, we re-applied the multi-period DID model to estimate the influence of big data on carbon emissions. The stepwise regression findings are delineated in Table 3. Notably, no matter how many control variables are added, the coefficient of the policy dummy variable consistently exhibits significant negativity. This reaffirms the inhibiting effect of big data on carbon emissions.

Table 3. PSM-DID retest results.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Did	−0.130 ***	−0.131 ***	−0.096 **	−0.095 **	−0.104 **	−0.096 **
	(0.037)	(0.037)	(0.038)	(0.037)	(0.038)	(0.037)
Constant	5.629 ***	5.370 ***	2.490 **	4.401 ***	4.476 ***	−0.430
	(0.042)	(0.130)	(0.789)	(1.208)	(1.210)	(1.322)
Control	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2120	2120	2120	2120	2120	2120
R-square	0.908	0.909	0.910	0.910	0.910	0.913

Note: *** $p < 0.01$, ** $p < 0.05$.

5.3. Robustness Tests

In order to ensure the reliability of the above estimation results, this study undertakes robustness tests through four distinct methodologies: parallel trend tests, the replacement of dependent variable, the placebo test, the exclusion of other policies and the endogeneity test.

5.3.1. Parallel Trend Test

In order to ensure the unbiasedness and validity of the estimation results, it is necessary to satisfy the parallel trend assumption before applying the DID model for policy evaluation, which means that the carbon emissions of the control group and the experimental group should maintain the same change trend as before the establishment of the NBD-CPZ. Based on Beck et al. (2010) [36], this study adopts the event study method to conduct the parallel trend test, and the model is set as follows:

$$\begin{aligned} Carbon_{it} &= \beta_0 + \beta_1 Policy_{it}^{-4} + \beta_2 Policy_{it}^{-3} + \beta_3 Policy_{it}^{-2} + \cdots \\ &+ \beta_7 Policy_{it}^{3} + \beta_8 Policy_{it}^{4} + \gamma X_{it} + \mu_i + \nu_t + \varepsilon_{it} \end{aligned} \quad (2)$$

where $Policy_{it}^{\pm j}$ is a dummy variable. When the experimental group i is in the year j before and after the policy shock, $policy_{it}^{\pm j}$ is set to 1, otherwise it remains 0, and its coefficient serves to depict the dynamic effect of big data on carbon emissions. Moreover, the measurement of the dependent variable and the control variables aligns with the basic regression. The model also incorporates the controls for city and year effects, ultimately remaining a two-way fixed effects model. The results of the parallel trend test are presented in Table 4. These results indicate that the coefficients are not significant within the first four years after the establishment of the NBD-CPZ. Conversely, during the latter four years following the NBD-CPZ's construction, the coefficients exhibit statistically significant negative values. This suggests that there are no substantial differences in carbon emission levels between the experimental and control groups, thereby passing the parallel trend test.

Table 4. Parallel trend test results.

Variables	*Carbon*
$Policy^{-4}$	−0.028
	(0.065)
$Policy^{-3}$	−0.032
	(0.058)
$Policy^{-2}$	0.007
	(0.058)
$Policy^{-1}$	Base period
Current	0.033
	(0.062)
$Policy^{1}$	−0.113 *
	(0.068)
$Policy^{2}$	−0.120 *
	(0.067)
$Policy^{3}$	−0.247 ***
	(0.080)
$Policy^{4}$	−0.338 **
	(0.168)

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5.3.2. Replacement of Dependent Variable

Drawing upon the research conducted by Li and Zhan (2022) [46], we substitute the dependent variable with carbon emission intensity. This intensity is determined by the ratio of carbon emissions to GDP. Subsequently, this adjusted variable is incorporated into the basic model for regression analysis. The outcomes of this analysis, presented in regression (1) in Table 5, reveal a coefficient of −0.074 for *Did*, significant at the 10% level. Notably, there exists no substantial disparity in the magnitude and significance of this coefficient compared to the results obtained from the basic regression. This consistency underscores the reliability of the conclusions derived from the basic regression analysis.

Table 5. Robustness tests.

Variables	(1)	(2)	(3)	(4)	(5)
Did	−0.072 **			−0.080 **	−2.265 ***
	(0.036)			(0.037)	(0.418)
Did1		0.078			
		(0.063)			
Did2			−0.039		
			(0.054)		
Did3				−0.072 **	
				(0.029)	
Constant	7.823 ***	0.215	−0.460	−0.500	
	(1.263)	(2.414)	(2.454)	(1.315)	
Control	Yes	Yes	Yes	Yes	Yes
City Fe	Yes	Yes	Yes	Yes	Yes
Year Fe	Yes	Yes	Yes	Yes	Yes
Kleibergen-Paap rk LM statistic					48.243
					[0.000]
Kleibergen-Paap rk Wald F statistic					56.736
					{16.38}
Observations	2162	2162	2136	2162	2155
R-square	0.834	0.618	0.618	0.912	

Note: [] is the *p*-value; { } is the critical value at the 10% level of the Stock-Yogo test. *** $p < 0.01$, ** $p < 0.05$.

5.3.3. Placebo Test

The placebo test, commonly employed in research, incorporates two approaches: the fictitious experimental group and the fictitious policy time. In the context of the fictitious experimental group, drawing upon existing literature, this study interchanges the experimental and control groups. Subsequently, the multi-period DID model is employed to re-evaluate the effect of big data on urban carbon emissions, depicted in regression (2) within Table 5. Notably, the coefficient of Did1 demonstrates statistical insignificance, suggesting that big data does not exert a significant influence on urban carbon emissions. Regarding the fictitious policy time, as the construction of a NBD-CPZ occurred solely in the Guizhou Province in 2015, this study excludes the Guizhou Province from the experimental group. Additionally, the approved construction time for the remaining experimental group is adjusted from 2016 to 2015. Subsequently, the DID model is employed for estimation, with the regression outcomes presented in regression (3) of Table 5. Similarly, the coefficients of Did2 exhibit insignificance. In summary, whether employing a fictitious experimental group or fictitious policy time, the impact of big data on carbon emissions is statistically insignificant, that is, the conclusions of this study are credible.

5.3.4. Exclusion of Other Policies

It is imperative to acknowledge that when evaluating the carbon impact of big data based on the policy shock of NBD-CPZs, the influence of other policy shocks on carbon emissions is unaddressed. This oversight potentially undermines the accurate estimation of the carbon effect of big data. Through a comprehensive examination of policies in the surrounding years of 2015 and 2016, alongside relevant studies, it is found that the "Broadband China" pilot policy significantly enhances carbon emission performance [47]. This finding suggests that the "Broadband China" pilot policy may affect the accuracy of the conclusions drawn in this study. Consequently, this study constructs a dummy variable for the "Broadband China" pilot policy and includes it in the basic model for estimation, with the results presented in regression (4) of Table 5. The results show that the coefficient of *Did* remains significantly negative at the 5% level, even after accounting for the dummy variable of the "Broadband China" pilot policy. Thus, the assertion that big data can substantially mitigate carbon emissions stands as robust and reliable.

5.3.5. Endogeneity Test

Since NBD-CPZs are not randomly established, this, in turn, leads to the endogeneity problem in policy estimation. For this reason, this study adopts the instrumental variable method to address the endogeneity problem. Referring to the study of Lin and Tan (2019) [48], the interaction term between terrain relief and year is used as an instrumental variable. The endogeneity test is shown in regression (5) of Table 5. The results show that the coefficient of Did remains significantly negative after accounting for endogeneity. In addition, the test results of both the Kleibergen-Paap rk LM statistic and the Kleibergen-Paap rk Wald F statistic confirm that the instrumental variables are reasonably valid. Therefore, the conclusion that big data can reduce urban carbon emissions still holds when endogeneity is excluded.

5.4. Mechanism Analysis

The previous theoretical analysis elucidates that big data reduces carbon emissions by promoting green innovation, optimizing energy structure, mitigating resource mismatch and improving public awareness of environmental protection. Acknowledging the substantial endogeneity inherent in the mediating effect model, this study refers to the study of Jiang (2022) [49] and focuses on the influence of the core explanatory variable on the mechanism variables. The ensuing econometric model, constructed accordingly, is presented below:

$$M_{it} = \alpha_0 + \alpha_1 Did_{it} + \gamma C_{it} + \mu_i + \omega_t + \varepsilon_{it} \quad (3)$$

where M symbolizes the mechanism variables, encompassing green innovation (*GI*), energy structure (*ES*), resource mismatch (*RM*) and public awareness of environmental protection (*Awareness*). The remaining variables maintain the same settings as in the fundamental regression analysis. Specifically, green innovation is quantified by the logarithm of green patent authorizations from previous years, while the energy structure is represented by the proportion of coal consumption in energy utilization. Moreover, for resource mismatch, the approach adopted in this study aligns with that delineated by Bai et al. (2018) [50], employing the capital mismatch index and labor mismatch index for measurement. Public awareness of environment protection is measured using the level of public concern for the environment in each region.

The results of the mechanism test are delineated in Table 6. In column regression (1), the regression analysis reveals the impact of big data on green innovation, with a coefficient of *Did* at 0.318, significant at the 5% level. This finding suggests that big data plays a pivotal role in diminishing information transfer costs, alleviating information asymmetry, ensuring efficiency across various phases of green innovation endeavors and fostering urban green innovation, thereby contributing to carbon emission reduction efforts. Moving to column (2), which presents the regression results of big data on energy structure, the coefficient of *Did* is -0.042, significant at the 1% level. This indicates that big data leads to a reduction in the proportion of coal in energy consumption, thereby optimizing the energy consumption structure and curbing carbon dioxide emissions at their source. Subsequent columns, (3) and (4), exhibit the regression outcomes of big data on capital mismatch and labor mismatch, respectively. Notably, a significantly negative coefficient of *Did* is observed solely in the case of capital mismatch. Thus, it becomes evident that the attenuating impact of big data on carbon emissions is chiefly achieved through the mitigation of capital mismatch. Column (5) shows the regression results of big data on public awareness of environmental protection and the coefficient of *Did* is significantly positive at the 1% level, indicating that big data can improve public awareness of environmental protection and then form low-carbon behavioral styles and help energy conservation and emission reduction. In summary, it is apparent that big data can effectively contribute to regional carbon emission reduction by promoting green innovation, optimizing energy structure, mitigating capital mismatch and improving public awareness of environmental protection.

Table 6. Mechanism tests.

Variables	(1)	(2)	(3)	(4)	(5)
	GI	***ES***	***RM-Capital***	***RM-Labor***	***Awareness***
Did	0.097 ***	−0.042 ***	−0.025 **	0.011	0.104 ***
	(0.031)	(0.010)	(0.012)	(0.011)	(0.026)
Constant	−5.544 ***	4.569 ***	1.991 ***	3.347 ***	−3.648 ***
	(1.236)	(0.404)	(0.550)	(0.475)	(1.098)
Control	Yes	Yes	Yes	Yes	Yes
City Fe	Yes	Yes	Yes	Yes	Yes
Year Fe	Yes	Yes	Yes	Yes	Yes
Observations	2162	2130	2123	2123	2144
R-square	0.966	0.780	0.903	0.850	0.934

Note: *** $p < 0.01$, ** $p < 0.05$.

5.5. Heterogeneity Analysis

5.5.1. Heterogeneity of Urban Digital Economy Level

This study, drawing upon the research of Han and Han (2023) [51], has established an evaluation system for digital economy levels, employing the entropy method to gauge the digital economy level of cities. Subsequently, the city samples were categorized into two groups based on their digital economy level: cities characterized by a high digital economy level and those with a low digital economy level determined by their annual median. The results of the heterogeneity analysis from the perspective of the urban digital economy level are shown in regressions (1)–(2) in Table 7. The regression results unveil a significant negative coefficient of *Did* within the cohort of cities exhibiting a high digital economy level at a 10% significance level. Conversely, the coefficients of *Did* in the low digital economy level city samples prove to be statistically insignificant, suggesting that big data's carbon reduction impact manifests primarily in cities boasting high digital economy levels. The reason for this is that data plays a pivotal role in the development of the digital economy. Contrasted with cities with low digital economy level, cities with high digital economy levels possess more sophisticated big data technology, enterprises carry out digital transformation earlier and the completion degree of the digital transformation of enterprises is also higher. These conditions foster the efficient and prompt realization of information integration and sharing, enhancing resource allocation efficiency, consequently leading to a reduction in carbon emissions.

Table 7. Heterogeneity analysis.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Digital Economy Level		**Resource Endowments**		**IPR Protection Intensity**		**Policy Impact Time**	
	High	**Low**	**Resource-Based City**	**Non-Resource-Based City**	**Strong**	**Weak**	**Guizhou**	**Other Pilot Cities**
Did	−0.135 *	−0.064	−0.018	−0.127 **	−0.251 **	0.008	−0.387 **	−0.065 *
	(0.071)	(0.059)	(0.059)	(0.048)	(0.077)	(0.051)	(0.122)	(0.038)
Constant	−0.135 *	−0.064	−0.018	−0.127 **	−0.251 **	0.008	−3.984 **	−1.028
	(0.071)	(0.059)	(0.059)	(0.048)	(0.077)	(0.051)	(1.623)	(1.336)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City Fe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1062	1100	900	1262	1055	1107	1599	2136
R-square	0.929	0.939	0.881	0.926	0.931	0.907	0.914	0.912

Note: ** $p < 0.05$, * $p < 0.1$.

5.5.2. Heterogeneity of Urban Resource Endowments

In this study, based on the National Plan for Sustainable Development of Resource-Based Cities (2013–2020) issued by China, the cities under study have been categorized

as either resource-based or non-resource-based. The results of heterogeneity analysis from the perspective of resource endowment are shown in regressions (3)–(4) in Table 7. Findings reveal that the estimated coefficient of *Did* is significantly negative at the 5% significance level in the regression involving non-resource-based cities, whereas in the regression encompassing resource-based cities, this coefficient is negative but lacks statistical significance. This discrepancy signifies substantial variations in the impact of big data on carbon emissions across cities with differing resource endowments. Specifically, big data reduces carbon emissions in non-resource-based cities, whereas it demonstrates no discernible effect on the carbon emissions of resource-based cities. Compared with resource-based cities, non-resource-based cities boast a more streamlined industrial structure and demonstrate a greater proclivity for innovation, thereby facilitating industrial transformation and upgrading after being empowered by big data. Furthermore, non-resource-based cities generally exhibit a higher overall level of economic development and marketization, and they pay more attention to environmental quality. The advancement of big data and other digital technologies enhances public engagement and enthusiasm for participating in environmental oversight, consequently forcing cities toward green transformation and reducing urban carbon emissions.

5.5.3. Heterogeneity of Intellectual Property Right Protection Intensity

In this study, based on the study of Hu et al. (2012) [52], the intensity of IPR protection is measured by the proportion of technology trade market turnover in the local GDP. Given that the National Bureau of Statistics of China exclusively divulges technology transaction market turnover at the provincial level, the intensity of IPR protection in prefecture-level cities is expressed by the proportion of the prefecture-level cities' GDP to the GDP of the province where it is located, multiplied by the intensity of IPR protection at the provincial level. Based on the median, the sample of cities is categorized into cities with strong IPR protection and cities with weak IPR protection. The tset results from the perspective of IPR protection intensity are shown in regression (5)–(6) in Table 7. As can be seen from the results, the coefficient of *Did* for cities with strong IPR protection is −0.251, which is significant at the 5% level, while the coefficient of *Did* for cities with weak IPR protection is statistically insignificant, which indicates that there is a significant difference in the impact of big data on carbon emissions under different IPR protection intensity, that is, under cities with strong IPR protection, big data is more capable of showing significant carbon emission reduction effects. In contrast to cities with weak IPR protection, those with strong IPR protection demonstrate a greater capacity to safeguard the legitimate rights and interests of innovators. This capability fosters a conducive environment for stimulating innovation enthusiasm among creators, thereby propelling green technological innovation and subsequently leading to decreased carbon emissions. In addition, the technological innovation effect of IPR protection can promote the empowerment of big data to various industries and accelerate the digital transformation of various industries, thereby reducing energy consumption and reducing carbon emissions from the source.

5.5.4. Heterogeneity of Policy Impact Time

Since the Guizhou Province and other pilot cities established their NBD-CPZs at different times, this study divides the sample of the experimental group into the Guizhou Province and other pilot cities according to the time of policy impacts, in order to explore whether the difference in the time of the policy impacts between Guizhou Province and other pilot cities causes a difference in the carbon-reducing effect of the NBD-CPZ. The results of heterogeneity analysis from the perspective of policy impact time are shown in regressions (7)–(8) in Table 7. The results show that the estimated coefficients of *Did* are significantly negative in both groups of regressions, but compared with the estimated coefficients of *Did* in regression (8), the estimated coefficients and significance of *Did* in regression (7) are significantly higher, which suggests that the carbon emission reduction effect of the NBD-CPZ in the Guizhou Province is significantly stronger than that of

the carbon emission reduction effect of the NBD-CPZ in other pilot cities. The Guizhou Province was the construction site of China's first NBD-CPZ partly due to the fact that the Guizhou Province started early in the big data industry and has a good foundation for the development of big data. Therefore, compared to other pilot cities, the Guizhou Province can more quickly promote the application of big data in various industries, help industrial transformation and upgrading, promote green innovation of enterprises and thus reduce carbon emissions to a greater extent. In addition, compared to other pilot cities, the Guizhou Province established its NBD-CPZ earlier, so it has more experience in policy implementation, better guarantees policy implementation, and helps carbon emission reduction.

6. Analysis of Spatial Effects

In this study, a spatial autocorrelation test of carbon emissions in Chinese cities was conducted, and the results are shown in Table 8. As can be seen from the results, the Moran's I value of each city from 2012 to 1016 are statistically insignificant, while those of each city from 2017 to 2019 are statistically significant, which indicates that the spatial correlation of the carbon emissions of each city is constantly strengthening. Therefore, it is necessary to further examine the spatial effect of the NBD-CPZ on urban carbon emissions.

Table 8. Spatial autocorrelation test of carbon emissions from 2012 to 2019.

Year	Moran's I	*p*-Value
2012	−0.017	0.680
2013	−0.026	0.484
2014	−0.026	0.495
2015	−0.042	0.240
2016	−0.047	0.182
2017	−0.074	0.030
2018	−0.060	0.085
2019	−0.058	0.093

Hence, there is a pressing need to investigate the spatial effect of the NBD-CPZ on urban carbon emissions. In view of this, this study initiates the construction of a spatial Durbin difference-in-difference model (SDM-DID). Through a series of rigorous tests, the specific form of the model is determined, with detailed results presented in Table 9. Notably, the results of the LM, Wald, LR, Hausman and two-way fixed effects tests significantly reject the original hypothesis. Consequently, it is justifiable for this study to adopt the SDM model, which incorporates controls for both city-fixed effects and time-fixed effects. In summary, the model constructed in this research is delineated below:

$$Carbon_{it} = \rho WCarbon_{it} + \beta Did_{it} + \theta WDid_{it} + \gamma X_{it} + \sigma WX_{it} + \mu_i + \nu_t + \varepsilon_{it} \quad (4)$$

where ρ represents the spatial autocorrelation coefficient, β signifies the impact coefficient of big data on local carbon emissions, θ denotes the coefficient reflecting the influence of big data on carbon emissions in neighboring regions, and γ symbolizes the coefficient of the control variables. Additionally, W represents the spatial weighting matrix, which is used to describe the spatial correlation degree between the city i and the city j. In this study, we adopt the inverse distance squared spatial weight matrix, as shown in Formula (5), where dij2 denotes the squared distance between distinct cities:

$$W = \begin{cases} 1/d_{ij}^2 \ (i \neq j) \\ 0 \ (i = j) \end{cases} \quad (5)$$

Table 9. Spatial model selection tests.

Type of Test	Statistical Results	*p*-Value
Lm_lag	3.247	0.072
Robust Lm_lag	0.312	0.576
Lm_error	146.640	0.000
Robust Lm_error	143.705	0.000
Wald_spatial lag	56.29	0.000
Wald_spatial error	71.99	0.000
Can SDM be degraded to SAR	16.31	0.012
Can SDM be degraded to SEM	15.49	0.017
Hausman test	120.48	0.000
Selection area fixed effects	236.97	0.000
Selection time fixed effect	2708.66	0.000

Regressions (1)–(3) in Table 10 show the regression results derived from three spatial models: SEM, SAR and SDM, respectively. Notably, the spatial autocorrelation coefficients, denoted by ρ or λ across the three models, exhibit statistically significant positivity. This observation underscores the pronounced spatial clustering tendencies within carbon emissions among cities, thereby reinforcing the inclusion of spatial factors within this study's framework. Furthermore, the coefficients of *Did* across three models manifest as significantly negative, thereby reaffirming the inhibitory impact of big data on carbon emissions within pilot cities. Upon scrutinizing the effect of decomposition findings from the SAR and SDM models, it is discerned that the coefficient of the spatial term of the policy dummy variable ($W \times Did$) registers as positive, yet statistically insignificant. This insignificance implies that the NBD-CPZ exerts no notable influence on carbon emissions within neighboring regions. One plausible explanation for this result may be that the NBD-CPZ has both siphon and spillover effects on neighboring regions.

Table 10. Results of spatial effects.

Variables	(1)	(2)	(3)
ρ or λ	0.131 ***	0.159 ***	0.166 ***
	(0.025)	(0.045)	(0.025)
Did	−0.085 ***	−0.086 ***	−0.082 **
	(0.032)	(0.032)	(0.033)
$W \times Did$			0.170
			(0.159)
	(0.030)		(0.154)
Control	Yes	Yes	Yes
Year Fe	Yes	Yes	Yes
City Fe	Yes	Yes	Yes
Observations	2248	2248	2248
Log-likelihood	−722.322	−722.733	−714.579
R-square	0.279	0.283	0.254

Note: *** $p < 0.01$, ** $p < 0.05$.

In order to further identify the spatial structure of the siphon effect and spillover effect in the NBD-CPZ, this study draws on the study of Sun and Yuan (2019) [53] to construct the spatial weight matrix in the form of inverse distance decay. To guarantee each city's inclusion of at least one neighboring counterpart, this study sets the initial geographic distance threshold at 400 km, with a subsequent increment of 100 km. The computation of the spatial decay matrix follows this formulation, wherein D represents the geographic distance threshold:

$$W_{\mathrm{d}} = \begin{cases} 1/d_{ij}^2, \ d_{ij} \geq D \\ 0, \ d_{ij} < D \end{cases} \quad (6)$$

Table 11 shows the spatial impact of the NBD-CPZ on carbon emissions at different geographic distances. The results delineate three discernible trends: initially, within the 400–500 km threshold, the spatial effect of the impact of the NBD-CPZ on carbon emissions is significantly positive, indicating that at this time, affected by the shadow of agglomeration, the NBD-CPZ has a siphon effect on the resources of the neighboring areas. Subsequently, within the 500–900 km threshold, the spatial effect of the impact of the NBD-CPZ on carbon emissions is significantly negative, signaling a significant spillover effect from the NBD-CPZ. Finally, as the geographic distance exceeds 900 km, the spatial impact of the NBD-CPZ becomes no longer significant. Thus, it becomes evident that the NBD-CPZ engenders both siphoning and spillover effects, and exhibits a certain spatial structure that corroborates Hypothesis 3.

Table 11. Spatial effects of the NBD-CPZ under different geographic distances.

Threshold (km)	Ratio	Standard Error
400	0.204 ***	0.044
500	−0.090 ***	0.016
600	−0.156 ***	0.024
700	−1.203 **	0.520
800	−2.165 *	1.219
900	−1.638	1.384
1000	0.179	1.040

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

7. Conclusions and Policy Recommendations

7.1. Conclusions

Drawing on the panel data from Chinese prefecture-level cities spanning the period from 2012 to 2019, this study leverages the establishment of the national big data comprehensive pilot zone (NBD-CPZ) as a quasi-natural experiment to empirically scrutinize the impact and mechanisms underlying big data's influence on carbon emissions. There are four key study findings. (1) Big data exerts a significant inhibitory effect on urban carbon emissions, a conclusion robustly validated through a battery of robustness tests. (2) Mechanism analysis underscores that big data operates along three distinct pathways to suppress urban carbon emissions: fostering green innovation, optimizing energy structure and mitigating capital mismatches improving public awareness of environmental protection. (3) An analysis of heterogeneity reveals disparities in the carbon emission reduction effects of big data across various city typologies. Notably, big data is more effective in reducing carbon emissions in cities with a high digital economic level, non-resource-based cities, cities with strong IPR protection and the Guizhou Province. (4) The analysis of spatial effects shows that within the span of 400–500 km, the NBD-CPZ manifests a siphon effect, resulting in increased carbon emissions. Conversely, in the range of 500–900 km, the NBD-CPZ manifests a spillover effect, resulting in reduced carbon emissions. Notably, beyond the 900 km, the spatial effect of the NBD-CPZ becomes statistically insignificant.

7.2. Policy Recommendations

Based on the findings of this study, we advance the following policy recommendations.

First, we should make every effort to promote the development of the big data industry. Empirical evidence shows that this can significantly reduce urban carbon emissions. Against the backdrop of the goal of carbon peaking and carbon neutrality, the comprehensive development of the big data sector holds profound strategic importance. Government intervention is pivotal in this effort, necessitating the establishment of a robust policy framework. Such a framework should not only include standardized management protocols for data resources, but also incorporate mechanisms pertaining to data resource rights. This approach serves to fortify support and provide clear guidance to the burgeoning big data industry, consequently fostering a conducive environment for its development.

Furthermore, given the exponential growth of the big data sector, there is a palpable surge in demand for skilled professionals among enterprises and research institutions. Consequently, the government should undertake initiatives to cultivate a substantial pool of big data talent through educational programs and specialized training, thereby injecting significant momentum into the industry's expansion.

Second, we should differentiate the focus point of big data carbon reduction. Given the variable impact of big data on carbon emissions across diverse cities, the strategic deployment of big data for emission reduction necessitates a comprehensive consideration of factors, such as the city's digital economy level, resource endowment and IPR protection intensity. Tailoring interventions to local contexts becomes essential. For instance, in cities with a low digital economy level, enhancing digital infrastructure, fostering information integration and sharing and optimizing resource allocation efficiency are imperative. Resource-based cities require the expedited integration of big data into traditional industries to facilitate their transition toward sustainability and reduce reliance on fossil fuels. Additionally, cities with weak IPR protection must strengthen the construction of an intellectual property rights protection system, elevate the costs of IPR infringement and then enhance the green innovation willingness of innovation subjects.

Third, we should explore the multidimensional carbon reduction pathways of big data. It is evident from mechanism tests that big data primarily serves to reduce urban carbon emissions through the promotion of green innovation, optimization of energy structure and mitigation of resource mismatch. Henceforth, it is imperative to promote the integration of big data with various industries while promoting its widespread use. This strategic endeavor aims to catalyze the digital transformation of industrial sectors, augment the proportion of clean energy in energy consumption patterns and improve resource mismatches. Additionally, sustained efforts are necessary to augment investment in research and development pertaining to green innovation, enhance enterprises' cognizance of green innovation imperatives and expedite the translation of green innovation outcomes into practical solutions, thereby effectuating the transition towards a low-carbon economy and society.

Fourth, regional cooperation aimed at reducing carbon emissions should be further promoted. From the analysis of spatial effects, it can be seen that the spatial spillover effect of the NBD-CPZ is significant in the surrounding 500–900 km. Consequently, municipalities ought to foster a collective commitment towards carbon emission reduction. Leveraging big data technology to surmount administrative barriers, enhancing inter-city communication and fostering cooperation is pivotal. Additionally, there is a pressing need to explore inter-regional collaborative carbon reduction mechanisms. Such endeavors not only bolster the spatial spillover effect of the NBD-CPZ, but also promote the low-carbon development of cities.

This study still has some limitations in exploring the relationship between big data and urban carbon emissions, mainly in two aspects. On the one hand, our study focuses only on the Chinese city level, ignoring other country samples. All countries around the world are responsible for combating global climate change. Therefore, subsequent studies can expand the scope of research and explore the relationship between big data and carbon emissions from an international perspective. On the other hand, our study is limited to the period of 2012–2019 and excludes city samples with a significant amount of missing data, which may lead to some bias in the results. If the missing data can be made up, the research conclusions will be more accurate.

Author Contributions: Conceptualization, Y.L. and H.C.; Methodology, Y.L.; Software, Y.L. and Z.L.; Validation, Y.L. and Z.L.; Formal analysis, Z.L. and H.C.; Investigation, H.C.; Resources, Z.L. and H.C.; Data curation, H.C.; Writing—original draft, H.C.; Writing—review & editing, X.C.; Visualization, H.C. and X.C.; Supervision, X.C.; Project administration, X.C.; Funding acquisition, X.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study. Requests to access the datasets should be directed to the Corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shi, Y.T.; Wang, X.D. Can the establishment of a big data comprehensive pilot zone drive the digital transformation of enterprises?—Empirical research based on quasi-natural experiments. *Stud. Sci. Sci.* **2024**, *42*, 1482–1492.
2. Sun, W.Z.; Mao, N.; Lan, F.; Wang, L. Policy empowerment, digital ecosystem and enterprise digital transformation: A quasi-natural experiment based on the national big data comprehensive experimental zone. *China Ind. Econ.* **2023**, *9*, 117–135.
3. Cui, J.J.; Yong, Y.L.; Wang, S.H. National big data comprehensive pilot zone and enterprise green technology innovation. *Explor. Econ. Issues* **2023**, *12*, 37–59.
4. Han, F.; Mao, X. Impact of intelligent transformation on the green innovation quality of Chinese enterprises: Evidence from corporate green patent citation data. *Appl. Econ.* **2024**, *56*, 5342–5359. [CrossRef]
5. Gong, Q.; Wang, X.; Tang, X. How can the development of digital economy empower green transformation and upgrading of the manufacturing industry?—A quasi-natural experiment based on the national big data comprehensive pilot zone in China. *Sustainability* **2023**, *15*, 8577. [CrossRef]
6. Yang, C. Digital economy drives regional industrial structure upgrading: Empirical evidence from China's comprehensive big data pilot zone policy. *PLoS ONE* **2023**, *18*, e0295609. [CrossRef]
7. Wei, J.Y.; Zhang, X.W. The role of big data in promoting green development: Based on the quasi-natural experiment of the big data experimental zone. *Int. J. Environ. Res. Public Health* **2023**, *20*, 4097. [CrossRef]
8. Gao, H.Y.; Xu, S.W.; Wang, M.Y. Will the National Big Data Comprehensive Pilot Zone improve total factor productivity of enterprises? *Energy Environ.* **2024**, 0958305X241241027. [CrossRef]
9. Wei, D.M.; Xu, Y.; Gu, N.H. Digital economy promote the high-quality economic development. *Sci. Res. Manag.* **2023**, *44*, 10–19.
10. Yang, Y.Q.; Wang, X.H. Digital empowerment and high-quality urban economic development: A quasi-natural experiment based on the national big data comprehensive pilot zone. *Inq. Econ. Issues* **2023**, 105–123.
11. Su, J.Q.; Tang, S.Y.; Zang, Y.Y. Does national-level comprehensive pilot zone of big data promote regional high-quality economic development: An empirical study based on natural experiment. *Mod. Financ. Econ. J. Tianjin Univ. Financ. Econ.* **2023**, *43*, 56–73.
12. Liu, D.; Wang, X.D.; Shi, Y.T. Impact of data elemants agglomeration on enterprise employment—Based on the dual perspective of employment scale and employment structure. *J. Shanxi Univ. Financ. Econ.* **2024**, *46*, 54–67.
13. Acemoglu, D.; Restrepo, P. The race between man and machine: Implications of technology for growth, factor shares, and employment. *Am. Econ. Rev.* **2018**, *108*, 1488–1542. [CrossRef]
14. Song, H.S.; Lu, L.Q. Big data and labor income share: Evidence from national big data comprehensive pilot zone. *Econ. Perspect.* **2024**, *1*, 111–128.
15. Peng, G.Z.; Zhou, X.L. Data elemant allocation and common Prosperity: A quasi-natural experiment based on national big data comprehensive pilot zone. *Inq. Into Econ. Issues* **2024**, *4*, 87–102.
16. Xu, X.C.; Ren, X.; Chang, Z.H. Big data and green development. *China Ind. Econ.* **2019**, *4*, 5–22.
17. Cheng, Y.J.; Duan, X. Can the digital economy promote urban haze reduction and carbon reduction? quasi-natural empirical analysis based on eight national big data pilot zones. *Soft Sci.* **2024**, *38*, 8–15.
18. Chang, H.L.; Jin, B.; Xue, F. Impact of big data strategy on carbon emissions from electricity consumption—A quasi-natural experiment from national big data comprehensive pilot zone. *Res. Econ. Manag.* **2023**, *44*, 93–109.
19. Wei, X.; Cheng, M.; Duan, K.; Kong, X. Effects of big data on PM2.5: A study based on double machine learning. *Land* **2024**, *13*, 327. [CrossRef]
20. Xu, W.X.; Zhou, J.P.; Liu, C.J. The impact of digital economy on urban carbon emissions: Based on the analysis of spatial effects. *Geogr. Res.* **2022**, *41*, 111–129.
21. Yang, G.Q.; Wang, H.S.; Fan, H.S.; Yue, Z.Y. Carbon reduction effect of digital economy: Theoretical analysis and empirical evidence. *China Ind. Econ.* **2023**, *5*, 80–98.
22. Salahuddin, M.; Alam, K. Internet usage, electricity consumption and economic growth in Australia: A time series evidence. *Telemat. Inform.* **2015**, *32*, 862–878. [CrossRef]
23. Hamdi, H.; Sbia, R.; Shahbaz, M. The nexus between electricity consumption and economic growth in Bahrain. *Econ. Model.* **2014**, *38*, 227–237. [CrossRef]

24. Hasseb, A.; Xia, E.J.; Saud, S.; Ahmad, A.; Khurshid, H. Does information and communication technologies improve environmental quality in the era of globalization: An empirical analysis. *Environ. Sci. Pollut. Res.* **2019**, *26*, 8594–8608. [CrossRef] [PubMed]
25. Lu, W.C. The impacts of information and communication technology, energy consumption, financial development, and economic growth on carbon dioxide emissions in 12 Asian countries. *Mitig. Adapt. Strateg. Glob. Change* **2018**, *23*, 1351–1365. [CrossRef]
26. Shen, Y.; Yang, Z.; Zhang, X. Impact of digital technology on carbon emissions: Evidence from Chinese cities. *Front. Ecol. Evol.* **2023**, *11*, 1166376. [CrossRef]
27. Zhao, Y.; Xin, L. Research on green innovation countermeasures of supporting the circular economy to green finance under big data. *J. Enterp. Inf. Manag.* **2021**, *35*, 1305–1322.
28. Shi, D.; Sun, G.L. The influence of the integration of digital economy and real economy on green innovation. *Reform* **2023**, *2*, 1–13.
29. Wu, J.; Zuidema, C.; Gugerell, K.; Roo, G. Mind the gap! Barriers and implementation deficiencies of energy policies at thelocal scale in urban China. *Energy Policy* **2017**, *106*, 201–211. [CrossRef]
30. Du, K.R.; Li, J.L. Towards a green world:how do green technology innovations affect total-factor carbon productivity. *Energy Policy* **2019**, *131*, 240–250. [CrossRef]
31. Avom, D.; Nkengfack, H.; Fotio, H.K.; Totouom, A. ICT and environmental quality in Sub-Saharan Africa: Effects and transmission channels. *Technol. Forecast. Soc. Change* **2020**, *155*, 120028. [CrossRef]
32. Zhang, Z.R.; He, J. The impact of digital economy development on carbon emissions—A quasi-natural experiment from the national big data comprehensive pilot area. *Inq. Into Econ. Issues* **2024**, *6*, 153–174.
33. Bai, J.H.; Wang, X.Y.; Bian, Y.C. The impact of internet development on factors allocation distortion. *J. Quant. Technol. Econ.* **2022**, *39*, 71–90.
34. Chen, X.D.; Yang, X.X. The impact of digital economic development on the upgrading of industrial structure: Based on the research of Gray relational entropy and dissipative structure theory. *Reform* **2021**, *3*, 26–39.
35. Pan, X.W.; Fu, W.L. Environmental information disclosure and regional air quality: A quasi-natural experiment based on the PM2.5 monitoring. *J. Financ. Econ.* **2022**, *48*, 110–124.
36. Beck, T.; Levine, R.; Levkov, A. Big bad banks? The winners and losers from bank deregulation in the United States. *J. Financ.* **2010**, *65*, 1637–1667. [CrossRef]
37. Zhang, H.; Feng, F. Does informal environmental regulation reduce carbon emissions?—Evidence form a quasi-natural experiment of environmental information disclosure. *Res. Econ. Manag.* **2020**, *41*, 62–80.
38. Yadava, V.S.; Yadavb, S.S.; Gupta, S.R.; Meena, R.S.; Lal, R.; Sheoran, N.S.; Jhariya, M.K. Carbon sequestration potential and CO_2 fluxes in a tropical forest ecosystem. *Ecol. Eng.* **2022**, *176*, 106541. [CrossRef]
39. Zhao, X.C.; Li, F.S.; Yan, Y.Z.; Zhang, Q. Biodiversity in urban green space: A bibliometric review on the current research field andits prospects. *Int. J. Environ. Res. Public Health* **2022**, *19*, 12544. [CrossRef]
40. Li, W.; Wang, Y. Optimization of Urban Road Green Belts under the Background of Carbon Peak Policy. *Sustainability* **2023**, *15*, 13140. [CrossRef]
41. Liu, J.; Li, M.; Ding, Y. Econometric analysis of the impact of the urban population size on carbon dioxide (CO_2) emissions in China. *Environ. Dev. Sustain.* **2021**, *23*, 18186–18203. [CrossRef]
42. Wang, G.X.; Wu, J.K. Study on industry agglomeration, urban scale and carbon emission. *J. Ind. Technol. Econ.* **2012**, *31*, 68–80.
43. Li, Z.X.; Li, N.Y.; Wen, H.W. Digital economy and environmental quality: Evidence from 217 cities in China. *Sustainability* **2021**, *13*, 8058–8070. [CrossRef]
44. Zhou, X.Y.; Zhou, D.Q.; Wang, Q.W. How information and communication technology drives carbon emissions: A sector-level analysis for China. *Energy Econ.* **2019**, *81*, 380–392. [CrossRef]
45. Zhang, S.; Li, J.; Jiang, B.; Guo, T. Government intervention, structural transformation, and carbon emissions: Evidence from China. *Int. J. Environ. Res. Public Health* **2023**, *20*, 1343. [CrossRef]
46. Li, S.; Zhan, Y. Carbon emission reduction effect of smart city construction from the perspective of industrial transformation and upgrading. *J. Shanghai Univ. Financ. Econ.* **2022**, *24*, 3–18+107.
47. Guo, J.; Wang, H. Innovative choices of emission-reduction strategies under digital empowerment: Quasi-natural experiment based on the "Broadband China" Pilot Policy. *Ind. Econ. Res.* **2022**, *4*, 101–113+142.
48. Lin, B.Q.; Tan, R.P. Economic agglomeration and green economy efficiency in China. *Econ. Res. J.* **2019**, *54*, 119–132.
49. Jiang, T. Mediating effects and moderating effects in causal inference. *China Ind. Econ.* **2022**, *5*, 100–120.
50. Bai, J.H.; Liu, Y.Y. Can outward foreign direct investment improve the resource misallocation of China. *China Ind. Econ.* **2018**, *1*, 60–78.
51. Han, Z.; Han, F.Q. Digital economy, fiscal decentralization and high-quality economic development: Empirical analysis based on panel data of 224 prefecture-level cities in China. *Inq. Into Econ. Issues* **2023**, *6*, 11–23.

52. Hu, K.; Wu, Q.; Hu, Y.M. The effect of intellectual property rights protection on technology innovation: Empirical analysis based on technology trading market and provincial panel data in China. *J. Financ. Econ.* **2023**, *38*, 15–25.
53. Sun, D.M.; Yuan, Y.J. Synergy innovation and regional industrial upgrading form the spatial spillover perspective. *Stat. Res.* **2019**, *36*, 100–114.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.